Using Schematic Models to Understand the Microscopic Basis for Inverted Solubility in γD-crystallin

Irem Altan,[†] Amir R. Khan,[‡] Susan James,[¶] Michelle K. Quinn,[¶] Jennifer J. McManus,[¶] and Patrick Charbonneau[*,†,§]

[†]*Department of Chemistry, Duke University, Durham, NC 27708, USA*

[‡]*School of Biochemistry and Immunology, Trinity College Dublin, Dublin, Ireland*

[¶]*Department of Chemistry, Maynooth University, Maynooth, Ireland*

[§]*Department of Physics, Duke University, Durham, NC 27708, USA*

E-mail: patrick.charbonneau@duke.edu

Abstract

Inverted solubility–melting a crystal by cooling–is observed in a handful of proteins, such as carbomonoxy hemoglobin and γD-crystallin. In human γD-crystallin, the phenomenon is associated with the mutation of the 23^{rd} residue, a proline, to a threonine, serine or valine. One proposed microscopic mechanism entails an increase in surface hydrophobicity upon mutagenesis. Recent crystal structures of a double mutant that includes the P23T mutation allow for a more careful investigation of this proposal. Here, we first measure the surface hydrophobicity of various mutant structures of γD-crystallin and discern no notable increase in hydrophobicity upon mutating the 23^{rd} residue. We then investigate the solubility inversion regime with a schematic patchy particle model that includes one of three variants of temperature-dependent patch energies: two of the hydrophobic effect, and one of a more generic nature. We conclude that while solubility inversion due to the hydrophobic effect may be possible, microscopic evidence to support it in γD-crystallin is weak. More generally, we find that solubility inversion requires a fine balance between patch strengths and their temperature-dependent component, which may explain why inverted solubility is not commonly observed in proteins. We also find that the temperature-dependent interaction has only a negligible impact on liquid-liquid phase boundaries of γD-crystallin, in line with previous experimental observations.

1 Introduction

Proteins can self-organize into a rich variety of superstructures,[1] such as crystals,[2] virus capsids,[3] disease-forming aggregates,[4] and biomaterials.[5] A key challenge is understanding how microscopic features of solvated proteins can give rise to such complex structures, and eventually to design systems that reliably assemble as such.[5–10] In this context, coarse-grained models are especially valuable, because they help both pinpoint and abstract the microscopic features that can reproduce the experimentally observed behavior. (Because simulating protein self-assembly typically requires hundreds to thousands of protein copies, which are them-

selves comprised of thousands of atoms, such models are also a computational necessity.[11–13]) For example, even relatively simple models of short-ranged,[14,15] anisotropic pair interactions largely recapitulate the phase behavior of globular proteins.[1,16,17] Understanding the assembly of some systems, however, requires coarse-grained models with a richer set of features, such as shape anisotropy for viral capsid and amyloid fiber-forming proteins.[1,3] Capturing certain aspects of protein crystallization, which is key to protein structure determination by diffraction methods,[18] can also require enriched patchy particle models.[2]

Proteins that exhibit atypical solution behaviors provide essential tests of our understanding of the physico-chemical processes that underlie their assembly. One such phenomenon is the decrease of protein solubility with increasing temperature, i.e., inverted crystal solubility, which is observed in a few proteins, such as some single mutants of γD-crystallin,[19,20] and the wild type carbomonoxy-hemoglobin C.[21] (The temperature invariant solubility of apoferritin is a limit case.[22]) Thermodynamically, inverted solubility suggests that as temperature increases, the Gibbs free energy of crystallization decreases, and hence that the crystal becomes increasingly more stable than the fluid. The phenomenon is often attributed to a large and positive entropy gain upon crystallization. Crystal formation is then possible even if the enthalpy of crystallization is non-negative.[21,23,24] Because the solute contribution to the change in entropy is typically negative, the solvent contribution is traditionally considered to be the key microscopic determinant of the phenomenon.[21,23,24]

The association of inverted solubility in proteins with the hydrophobic effect also comes from our understanding of the aqueous solvation of hydrocarbons, which presents an analogous anomaly.[25] The minimal Muller-Lee-Graziano (MLG) model for this effect considers water as being in one of four states: disordered shell (ds), ordered shell (os), disordered bulk (db), and ordered bulk (ob).[26,27] It was used by Shiryayev et al. to estimate the phase diagram of model globular proteins with isotropic interactions assumed to be driven exclusively by hydrophobic interactions.[23] Although the resulting phase behavior does present an inverted solubility regime, it is unclear whether this feature would persist for more realistic

protein models, with a complex surface mosaic of hydrophilic and hydrophobic interactions. In other words, while the hydrophobic scenario for solubility inversion in proteins is thermodynamically sound, microscopic evidence for it remains limited. The generality of the underlying physical arguments is also seemingly incompatible with the relatively rare occurrence of inverted solubility in experiments.

Here, we examine this microscopic scenario in the context of a double mutant (R36S+P23T) of the human γD-crystallin, which forms two competing crystals: a normal solubility structure (DBN, PDB[28] ID: 6ETA) and an inverted solubility structure (DBI, PDB ID: 6ETC).[29,30] The solubility inversion is here most likely associated with the mutation in the 23$^{\text{rd}}$ residue because the single mutants P23T, P23S, and P23V also exhibit inverted solubility,[31] albeit without generating diffraction quality crystals. In earlier work, we have parameterized a patchy model for the R36S+P23T mutant and obtained a solubility inversion regime by completely deactivating the patch containing the 23$^{\text{rd}}$ residue at low temperatures.[30] Interestingly, the DBI crystal does not present any obvious structural feature that could explain this effect, other than the formation of a hydrogen bond through the 23$^{\text{rd}}$ residue. Here, we critically evaluate three different temperature-dependent interaction potentials: the generic model we previously considered, and two that explicitly model the hydrophobic scenario. We use these models to test the hydrophobic scenario as well as the robustness of the inverted solubility regime with respect to model parameters. We thus attempt to elucidate why inverted solubility is not more commonly observed. We further explore the relationship between the liquid-liquid critical point and the solubility curve, which has been experimentally studied for some of these systems.[31] The plan for the rest of this paper is as follows. We first survey γD-crystallin crystal structures available in literature to determine whether an increase in surface hydrophobicity can be discerned upon introducing the solubility inverting mutations (Sec. 2). We then introduce a patchy protein model for these proteins (Sec. 3.1) along with the different temperature-dependent patch models (Sec. 3.2), and the methods used to determine solubility lines (Sec. 3.3). Sections 4.1 and 4.2 provide a detailed analysis

of these patchy models, and we conclude with proposals for further discerning experiments in Sec. 5.

2 Experimental Context for Hydrophobicity

As a first consideration of the reasonableness of the hydrophobicity scenario, we evaluate the surface hydrophobicity of various human γD-crystallin crystal structures. Were the P23T mutation to consistently increase surface hydrophobicity, one would infer the existence of strong evidence for the decrease in protein solubility upon mutagenesis to be driven by the hydrophobic effect. By studying the relative binding propensity of two dyes known to bind hydrophobic surfaces, Pande et al. indeed have previously inferred that P23T, P23S, and P23V mutants of human γD-crystallin do present a higher surface hydrophobicity than the wild type (WT) protein.[20] Here, in order to test the robustness and microscopic validity of this interpretation, we consider different scales that quantify hydrophobicity at the amino acid level. More specifically, we compute an average hydrophobicity of solvent-exposed residues[32] weighted by their solvent accessible surface area (SASA),[33] for five different scales: the grand average of hydropathy (GRAVY),[34] as well as the scales of Wimley and White (**ww**),[35] Hessa et al. (**hh**),[36] Moon and Fleming (**mf**),[37] and Zhao and London (also known as transmembrane tendency, **tt**).[38] Each of these scales assigns a hydrophobicity index to each residue type; all but **hh** and **mf** assign positive values to hydrophobic residues.

We compute hydrophobicity for three sets, S, of amino acids: (i) the entire protein surface, (ii) the surface of its N-terminus, i.e., the first 82 residues (including the solubility inverting 23$^{\text{rd}}$ residue), and (iii) the surface residues in the DBI contact that includes the 23$^{\text{rd}}$ residue (Patch 4 as per Sec. 3).[30] The hydrophobicity, H_ζ, for a given scale ζ is then obtained as

$$H_\zeta = \frac{\sum\limits_{i \in S} f_\zeta(i) A(i)}{\sum\limits_{i \in S} A(i)}, \tag{1}$$

where $f_\zeta(i)$ is the hydrophobicity index for residue i, and $A(i)$ is its SASA. We specifically consider: WT (PDB ID: 1HK0[39]), the P23T single mutant (PDB ID: 4JGF[40]), the R36S single mutant (PDB ID: 2G98[41]), the R58H single mutant (PDB ID: 1H4A[39]), DBI (PDB ID: 6ETC[30]), and DBN (PDB ID: 6ETA[30]). Of these, only WT, R36S, and R58H do not have a mutation at the 23$^{\text{rd}}$ residue. Note that missing residues are completed using Modeller[42] within Chimera,[43] and all crystal water molecules are removed prior to running this analysis. In order to estimate the error on these measured hydrophobicities, 100 configurations per crystal structure are created by perturbing particle coordinates by a random number selected from a Gaussian distribution with a standard deviation corresponding to the coordinate error specified in the PDB file. Two assumptions are made in estimating these error bars. First, the coordinate error reported in the PDB entry is assumed to be distributed uniformly and isotropically across all protein atoms. In reality, certain domains or residues in proteins are more mobile and thus harder to resolve by X-ray diffraction than others, but finer, residue-level information is not available. This assumption thus overestimates the error in more localized parts of the protein and underestimates the error in more mobile parts. Second, the refined structures do not precisely capture the actual protein structure, as suggested by R_{free} values ranging from 0.174 to as high as 0.284, hence possibly creating artificial hydrophobicity differences between different mutants, or, conversely, underestimating them.

The resulting hydrophobicity estimates are shown in Fig. 1. All measurements suggest that the entire protein, the N-terminus, and Patch 4 are overall hydrophilic, which is consistent with the fact that the protein is soluble in water. A more careful comparison is thus needed in order to determine whether certain substructures are more hydrophobic than others. We first compare the DBN and DBI structures, which are obtained from the same

double mutant, R36S+P23T, and which are structurally very similar.[30] As expected, nearly all measurements for DBN and DBI overlap within their 95% confidence intervals. The only exceptions are the hydrophobicities of Patch 4 measured by the GRAVY and **mf** scales. The latter likely results from **mf** uniquely classifying prolines as hydrophobic. This discrepancy could then amplify the minute difference in surface exposure of Patch 4 prolines between DBI and DBN. A similar argument could be made about GRAVY, as tyrosine is considered to be the most hydrophilic residue on the GRAVY scale.

Overall, the N-terminus is the most hydrophobic region in nearly all scales and for all structures. However, other observations are not consistent across scales. In particular, a number of nonmonotonicities can be observed. For instance, Patch 4 is more hydrophobic in DBI than in WT for the GRAVY, **hh**, and **mf** scales, but the **ww** and **tt** scales present no discernible difference. Similarly, Patch 4 is more hydrophobic in R36S than in WT for the **mf** scale, but the reverse is true for **hh**. These discrepancies reflect the different ordering of residues on different scales. For instance, GRAVY, which is calculated from experimental measurements of transfer free energies from water to water vapor, tends to assign aromatic side chains lower hydrophobicities than the other four scales, which instead consider the tendency of residues to transfer from bulk water to a lipid bilayer, a measurement prone to higher experimental uncertainty.[35]

Interestingly, the N-terminus of the P23T mutant is the least hydrophobic structure for the GRAVY and **mf** scales. This trend, however, disappears when only Patch 4 residues are considered. Patch 4, which controls solubility inversion, is actually *less* hydrophobic than the overall N-terminus or the entire protein, except on the **mf** scale. Only for this last scale is Patch 4 clearly more hydrophobic. A similar inconsistency is observed for Patch 4 of DBI, which is more hydrophobic than the other proteins for GRAVY and **mf**, but for these two scales P23T and DBN are not discernibly more hydrophobic than the structures without the mutation in the 23$^{\text{rd}}$ residue.

In summary, in none of the hydrophobicity scales do the structures with the (solubility-

inverting) P23T mutation have a statistically and consistently higher hydrophobicity than those without. P23T mutations even result in *lower* hydrophobicity estimates on some scales. While these results are subject to errors from the crystal structure accuracy, as well as the imperfections of the hydrophobicity scales themselves, a microscopic change to the protein surface that could putatively underlie the inversion of solubility nonetheless remains elusive. Because hydrophobicity scales are but an indirect measure of protein-water interactions (and thus of protein-protein interactions), however, other, more detailed approaches could be more revealing.

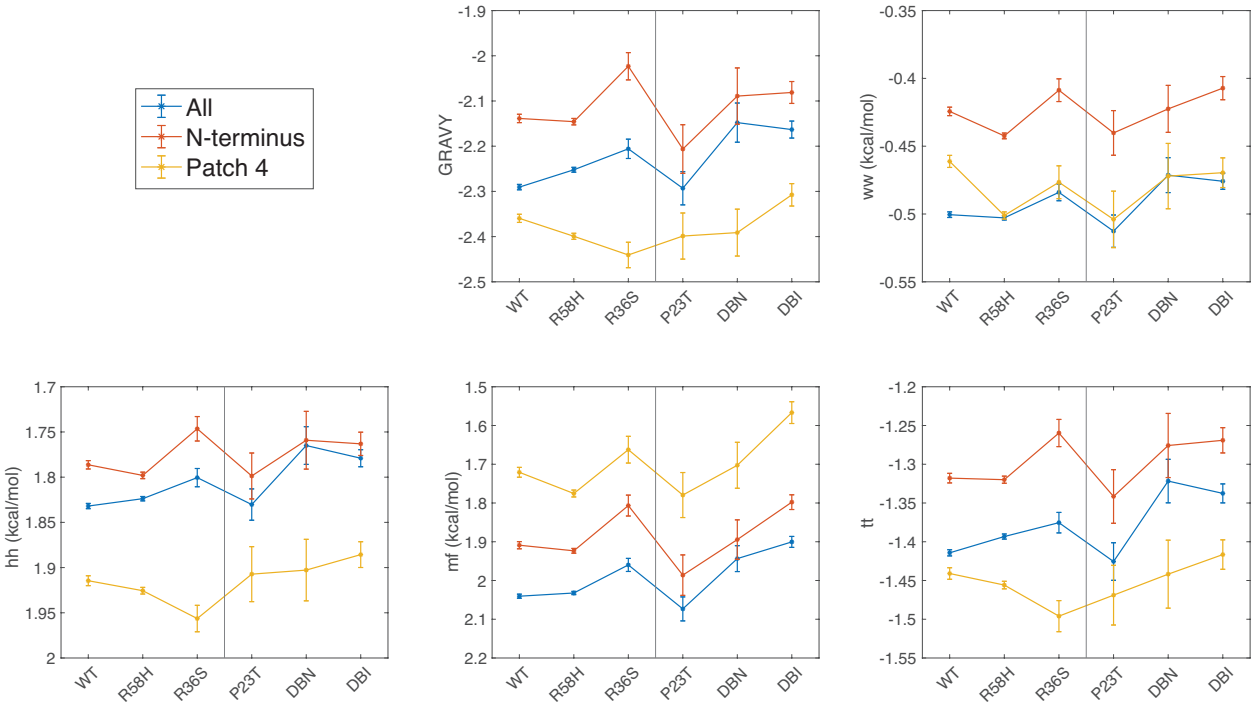

Figure 1: Hydrophobicity estimates for different crystal structures of single and double mutants of human γD-crystallin. Proteins to the left of the black vertical line exhibit normal solubility, and those to the right exhibit inverted solubility. Error bars denote 95% confidence intervals. Lines connecting the points are but a guide to the eye. Note the flipped scales for the **hh** and **mf** scales, in which lower values denote higher hydrophobicity, by contrast to the other scales. Structures with the P23T mutation do not systematically present a higher hydrophobicity, which is inconsistent with the hydrophobicity scenario.

3 Theory and Computational Details

Because a clear enhancement of hydrophobicity cannot be detected directly in mutants with inverted solubility, we next consider the thermodynamics of patchy models that incorporate various temperature-dependent patch energies. A schematic model of the double mutant of human γD-crystallin was previously studied in Ref. 30, and is here first modified to consider the hydrophobic scenario and then perturbed to evaluate the robustness of its inverted solubility regime.

3.1 Patchy Model

The schematic model consists of hard particles with attractive patches

$$u(r_{ij}, \Omega_i, \Omega_j) = u_{\mathrm{HS}}(r_{ij}) + \sum_{a,b}^{n} u_{ab}(r_{ij}, \Omega_i, \Omega_j), \tag{2}$$

where r_{ij} is the distance between particles i and j, Ω denotes the particle orientation, and $u_{\mathrm{HS}}(r_{ij})$ is the hard sphere potential for particles of diameter σ. The sum runs over all patch pairs, with n the total number of patches. The second contribution, u_{ab}, is further broken down into radial and orientational parts

$$u_{ab} = v_{ab}(r_{ij}) f_{ab}(\Omega_i, \Omega_j). \tag{3}$$

The radial part, v_{ab}, is a square-well interaction

$$v_{ab}(r_{ij}) = \begin{cases} -\varepsilon_{ab}(T), & \sigma < r_{ij} < \lambda_a + \lambda_b \\ 0, & \text{otherwise} \end{cases}, \tag{4}$$

with interaction ranges λ_a and λ_b of patches a and b, respectively, and with either constant or temperature-dependent patch energy $-\varepsilon_{ab}(T)$. The orientational part

$$f_{ab} = \begin{cases} 1, & \theta_{a,ij} \leq \delta_a \text{ and } \theta_{b,ij} \leq \delta_b \\ 0, & \text{otherwise} \end{cases} \times \begin{cases} 1, & \psi_{ij} \in [\varphi_{ab} - \Delta\varphi_{ab}, \varphi_{ab} + \Delta\varphi_{ab}] \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

contains two contributions. The first ensures that the relative particle orientation enables them to interact with δ_a and δ_b the angular width for patches a and b, respectively (Fig. 2a). The second limits the range $\varphi_{ab} \pm \Delta\varphi_{ab}$ of dihedral angles ψ_{ij} allowed for each pair (Fig. 2b), with $\theta_{a,ij}$ the angle between the vector defining the location of patch a and the vector that connects the centers of particles i and j, and $\theta_{b,ij}$ similarly for patch b.

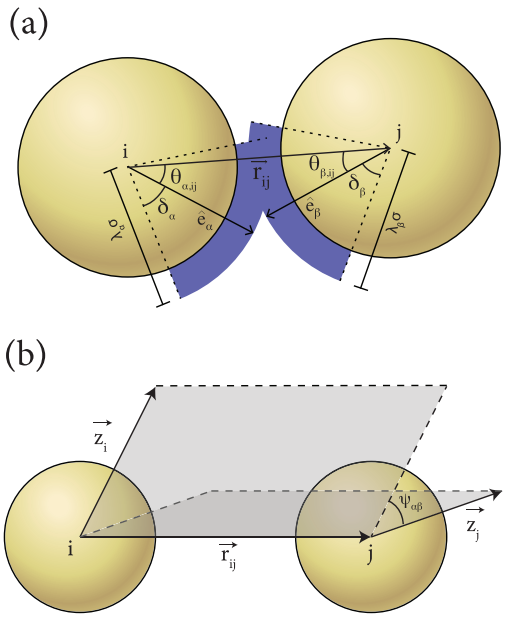

Figure 2: For two patches to interact, the relative particle orientation should satisfy the following. (a) The angle between the vector joining particles i and j, $\mathbf{r}_{ij}$, and the patch vectors $\hat{\mathbf{e}}_\alpha$ and $\hat{\mathbf{e}}_\beta$ should be less than δ_α and δ_β, respectively. (b) The dihedral angle between two particles, which is defined as the angle between two planes defined by the vectors $(\mathbf{z}_i, \mathbf{r}_{ij})$ and $(\mathbf{z}_j, -\mathbf{r}_{ij})$, should be within the range $\varphi_{ab} \pm \Delta\varphi_{ab}$. The reference vector $\mathbf{z}$ is chosen such that its orientation relative to the patches is identical for all particles.

This model is parameterized such that each patch corresponds to a crystal contact in

either the DBI or DBN crystal structure. This choice assumes that these surface patches are most chemically relevant for crystal formation, which is reasonable for such a small protein and is consistent with earlier studies of protein crystallization.[44,45] We then obtain five patches for DBI – labeled with Arabic numerals – and five patches for DBN – labeled with Roman numerals. Because Patch 4 of DBI contains the 23$^{\mathrm{rd}}$ residue, which is associated with the inverted solubility regime, this patch is taken to be temperature dependent (see Sec. 3.2); other patches are assigned a constant energy. Patch energies and interaction ranges were previously extracted from all-atom molecular dynamics simulations,[46] using umbrella sampling.[47] The resulting patchy particle model is sketched in Fig. 3, and the geometry details are given in the Supporting Information (Sec. S2). Note that the resulting effective single-component system model coarse-grains the role of solvent and ions in the crystallization cocktail. In what follows, unless otherwise specified, energies are reported in units of $k_{\mathrm{B}}T_{\mathrm{ref}}$, where $T_{\mathrm{ref}} = 277\mathrm{K}$ is the temperature at which DBN was crystallized experimentally, and distances are reported in units of the particle diameter σ, which here is taken to be 2.54 nm.

It is important to highlight that this protocol presents a number of limitations, including inaccuracies of the protein force field[48] and of the water model,[49] as well as the crudeness of representing potentials of mean force as square well interactions and proteins as spheres. In addition, determining the potential of mean force for each crystal contact is a computationally challenging task, and the sampling of each umbrella window might incompletely explore some of the protein conformational changes, such as loop motion.[50] On the whole, this approach likely yields estimates of protein-protein interactions that are at best within 10 to 50% of the association free energy. The associated phase diagram should therefore at least yield qualitatively, albeit not quantitatively, correct predictions.

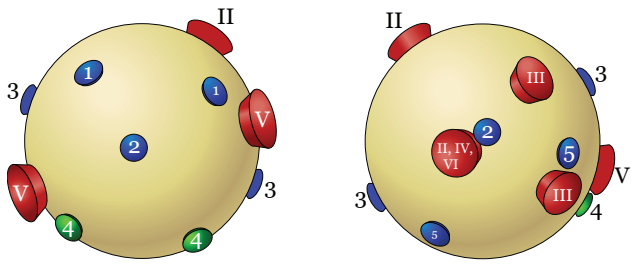

Figure 3: Front and back views of the patchy particle model. Blue and green patches are derived from DBI, and red patches from DBN. Patch 4 (green) contains the mutation associated with solubility inversion.

3.2 Inverted Solubility Models

In order to represent the microscopic origin of the inverted solubility, we consider three models for the temperature-dependence of the Patch 4 interaction energy: the MLG model, the Wentzel-Gunton model, and the temperature-(de)activated patchy model. Note that the parameters explicitly defined in these models are discussed in this section, while the free parameters are left for Sec. 4.1

MLG model– In this model, each of four water states is assigned a (relative) degeneracy, q, and an energy E. Degeneracies are ordered $q_{\mathrm{ds}} > q_{\mathrm{db}} > q_{\mathrm{ob}} > q_{\mathrm{os}}$. The last inequality follows from the hydration shell allowing the formation of only hydrogen bonds between water molecules and not to the hydrophobic solute. The higher degeneracy of the disordered shell compared to the disordered bulk follows from the additional orientational constraints in the former compared to the latter. Because only relative information about the degeneracies is needed, the estimates proposed in Ref. 51 here suffice: $q_{\mathrm{ob}} = 1.5$, $q_{\mathrm{db}} = 30$, $q_{\mathrm{os}} = 1$, and $q_{\mathrm{ds}} = 48$. Meanwhile, the energies are ordered $E_{\mathrm{ds}} > E_{\mathrm{db}} > E_{\mathrm{ob}} > E_{\mathrm{os}}$. The ordered shell is expected to have a lower energy than the ordered bulk state, because hydrogen bonds that form via tangentially oriented water molecules tend to be stronger than radially oriented ones; the disordered shell is expected to have a higher energy than the disordered bulk because replacing the solute with water molecules slightly increases the number of hydrogen bonds. Because energy values used by Ref. 25 are on an arbitrary scale, which is incompatible with the specific energy scale of our patchy model, we here use the values reported by Silverstein

et al. for the Mercedes-Benz model of water.[51,52] Posing that the energy of the ordered bulk is about one hydrogen bond, $E_{\rm ob} = -5.82\ k_{\rm B}T_{\rm ref}$,[53] the other three states have: $E_{\rm db} = -1.69\ k_{\rm B}T_{\rm ref}$, $E_{\rm os} = -5.90\ k_{\rm B}T_{\rm ref}$, and $E_{\rm ds} = -0.56\ k_{\rm B}T_{\rm ref}$. The energy and entropy per water molecule in the shell and are then given as[23]

$$E_{\rm s} = \frac{E_{\rm os} + E_{\rm ds}e^{-\beta(E_{\rm ds}-E_{\rm os})}}{1 + e^{-\beta(E_{\rm ds}-E_{\rm os})}} \tag{6}$$

$$E_{\rm b} = \frac{E_{\rm ob} + E_{\rm db}e^{-\beta(E_{\rm db}-E_{\rm ob})}}{1 + e^{-\beta(E_{\rm db}-E_{\rm ob})}} \tag{7}$$

and

$$s_{\rm s}/k_B = \log\left(\frac{q_{\rm os} + q_{\rm ds}e^{-\beta(E_{\rm ds}-E_{\rm os})}}{1 + e^{-\beta(E_{\rm ds}-E_{\rm os})}}\right) \tag{8}$$

$$s_{\rm b}/k_B = \log\left(\frac{q_{\rm ob} + q_{\rm db}e^{-\beta(E_{\rm db}-E_{\rm ob})}}{1 + e^{-\beta(E_{\rm db}-E_{\rm ob})}}\right). \tag{9}$$

The change in energy and entropy upon moving one water molecule from the bulk to the solvation shell of the protein are then simply $\varepsilon_{\rm w} = E_{\rm s} - E_{\rm b}$ and $\Delta s_{\rm w} = s_{\rm s} - s_{\rm b}$, respectively.

With this formulation the energy of Patch 4 is given by

$$\varepsilon_4' = \varepsilon_4 + n_{\rm w}\Delta\varepsilon(\beta) \tag{10}$$

where we have defined $\Delta\varepsilon(\beta) = 2(\varepsilon_{\rm w} - \Delta s_{\rm w}/\beta)$, and $n_{\rm w}$ is the number of water molecules in the solvation shell around contact i. Note that because patch parameters are measured at $\beta_{\rm ref} = 1$, parameters need to be tuned such that $\varepsilon_4'(\beta = 1) = \varepsilon_4$, and hence $\varepsilon_4' = (\varepsilon_4 - \Delta\varepsilon(1)n_{\rm w}) + \Delta\varepsilon(\beta)n_{\rm w}$. Note also that the temperature scale for the MLG model cannot be changed arbitrarily by changing $\beta_{\rm ref}$, because its parameters already set the range of temperatures within which the hydrophobic effect changes the free energy of crystallization.

Wentzel-Gunton Model– Wentzel and Gunton proposed a simplified version of the MLG

model in order to consider the phase behavior of particles with anisotropic interactions using Wertheim's theory.[24,54–56] This simple model assigns a linear temperature dependence for the patch energies

$$-\varepsilon_4' = -\varepsilon_4 - 2\varepsilon_{\mathrm{w}} + \frac{2}{\beta}\Delta s_{\mathrm{w}}, \tag{11}$$

where $-\varepsilon_{\mathrm{w}}$ and $-\Delta s_{\mathrm{w}}$ are free parameters that account for the change in energy and in entropy, respectively, due to the displacement of water upon contact association. Patch energies should equal those of the original model at β_{ref}, where the model was parameterized. This choice here suffices to set the overall temperature scale, because ε_{w} and Δs_{w} are arbitrary. Fixing ε_{w}, such that $\varepsilon_4'(\beta = \beta_{\mathrm{ref}}) = \varepsilon_4$, thus results in $\varepsilon_4' = \varepsilon_4 + 2\Delta s_{\mathrm{w}}(\frac{1}{\beta_{\mathrm{ref}}} - \frac{1}{\beta})$.

Temperature-(de)activated Patchy Model– de Las Heras and de Gama[57] proposed a model for patch (de)activation with temperature inspired by DNA-grafted colloids, which lose their attractive patches above the DNA melting temperature,[58] Although this model does not correspond to a specific microscopic scenario in proteins, it can nevertheless be construed as a simple and elegant way to (de)activate a patch. The temperature dependence of the interaction is then

$$\varepsilon_4'(T) = \frac{\varepsilon_4}{2}\left[1 + \tanh\left(\frac{T - T_{\mathrm{a}}}{\tau}\right)\right], \tag{12}$$

where T_{a} is the deactivation temperature, τ controls the sharpness of that deactivation. For this model, Patch 4 is deactivated below T_{a}.

3.3 Crystal Solubility Determination

Solubility lines are determined by first calculating the fluid and crystal chemical potentials, and then identifying the coexistence points at the intersection of these curves at fixed tem-

perature and pressure. For both DBI and DBN, experimental solubilities correspond to protein volume fractions of $\phi = 10^{-3}$ or lower.[29]

At such low densities simple local Monte Carlo (or molecular dynamics) sampling of the fluid phase is computationally inefficient, because transport is relatively slow. While this problem can be alleviated with advanced sampling methods such as aggregation volume bias Monte Carlo[59] and event chain Monte Carlo,[60] we here instead estimate the fluid properties from the second virial coefficient, B_2, which is calculated as in Ref. 61 (see SI Sec. S3[62]). Because the patch energies are high, B_2 can become very large and negative at low temperatures, but the protein density remains sufficiently low for $|B_2\rho| \ll 1$ in the regime of interest. In order to confirm that the third virial coefficient, B_3, can safely be neglected, we bound its value by noting that triply-bonded triplets of particles cannot form. The dominant contribution to B_3 thus comes from doubly-bonded triplets and scales as B_2^2, hence $|B_3|\rho^2 \lesssim |B_2|^2\rho^2 \ll 1$ in the regime of interest. Its contribution to the fluid chemical potential, μ_f, is therefore negligible, and so are higher-order corrections, thus justifying this theoretical expediency. The fluid equation of state and chemical potential can then be written as

$$\frac{\beta p}{\rho} = 1 + B_2\rho, \tag{13}$$

$$\beta\mu_\mathrm{f} = \beta\mu^\mathrm{id} + 2B_2\rho = \log\Lambda^3\rho + 2B_2\rho, \tag{14}$$

where $\beta\mu^\mathrm{id} = \log\Lambda^3\rho$ is the chemical potential of the ideal gas, and the thermal de Broglie wavelength Λ is set to unity, without loss of generality. With this formulation, we have

$$\rho = \frac{-1 + \sqrt{1 + 4B_2\beta p}}{2B_2}. \tag{15}$$

Note that by thermodynamical stability, $\beta\mu_{\rm f}$ must decrease with decreasing pressure, and our estimate is consistent in this respect. If B_2 is positive, $\partial\beta\mu_{\rm f}/\partial\rho$ is also positive; if B_2 is negative, $\partial\beta\mu_{\rm f}/\partial\rho > 0$ for $\rho < -B_2/2$, which is always true.

The crystal free energy at a given pressure and temperature is calculated using numerical simulations (see SI Sec. S3) with the Frenkel-Ladd method,[63] which involves thermodynamically integrating from an Einstein crystal. From this reference free energy, thermodynamic integration along an isobar provides the crystal chemical potential, $\mu_{\rm x}$, at different temperatures,

$$\beta\mu_{\rm x}(\beta, p) = \beta_0\mu_{\rm x}(\beta_0, p) + \int_{\beta_0}^{\beta} \frac{\langle H(\beta')\rangle}{N} d\beta' + \int_{\beta_0}^{\beta} \beta' \frac{\langle dU/d\beta'\rangle}{N} d\beta', \tag{16}$$

where $\langle H\rangle = p\langle V\rangle + \langle U\rangle$ is the enthalpy and $\langle\cdot\rangle$ denotes thermal averaging. Because of the highly constrained geometry of the patchy models, both crystals are almost incompressible. As a result, $\langle V\rangle$ is essentially independent of temperature. To high accuracy, we can thus write

$$\int_{\beta_0}^{\beta} \frac{\langle H(\beta')\rangle}{N} d\beta' \approx \frac{1}{N} \int_{\beta_0}^{\beta} \frac{\langle U(\beta')\rangle}{N} d\beta' + \frac{p}{\rho}(\beta - \beta_0), \tag{17}$$

and at sufficiently low pressures the second term on the right hand side is also negligible.

We further approximate that all the crystal bonds are active, and hence $\langle U(\beta)\rangle \approx U_0(\beta)$, where $U_0(\beta)$ is the ground state energy, and $\langle dU/d\beta\rangle \approx dU_0/d\beta$. While this last approx-

imation is generally quite good, it is overly crude in the patch deactivation regime, where the patch energy decreases rapidly around β_{a}, and vanishes when temperature is reduced further. As a result, $\langle dU/d\beta \rangle \ll dU_0/d\beta$, which can result in a significant correction to $\beta\mu_{\mathrm{x}}$ (see Fig. 4a). In the Wentzel-Gunton model, the patch similarly becomes non-attractive for $\beta > \beta_{\mathrm{ref}}$, and upon further lowering the temperature, it eventually becomes repulsive. The topology of the DBI crystal then changes and the energy of the crystal once again becomes temperature-independent, which leads to a bending of the evolution of the chemical potential with temperature (Fig. 4b). In both cases, however, the DBI solubility curve is unaffected, because these changes occur in a region where DBI is metastable with respect to DBN. Taking $\langle dU/d\beta \rangle \approx dU_0/d\beta$ is thus reasonable for our purposes.

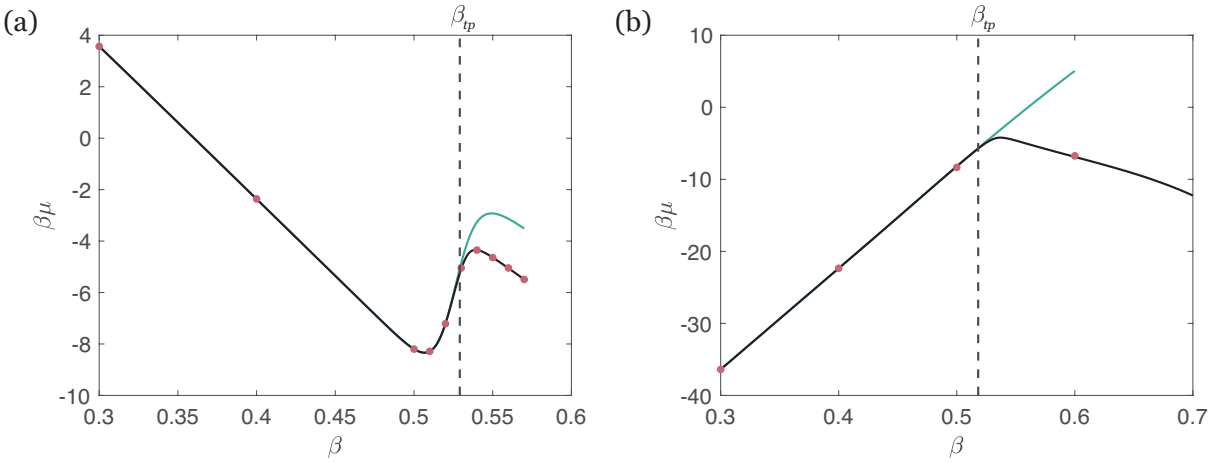

Figure 4: Calculated $\beta\mu_{\mathrm{x}}$ using thermodynamic integration starting from $\beta = 0.3$. The simulation data (black) fully matches the individual Einstein crystal simulations (red data points). Estimates of $\beta\mu_{\mathrm{x}}$ (blue) become significantly flawed at low temperatures, but because this regime is beyond the triple point, β_{tp} (dashed line), the DBI solubility line is unaffected. (a) Patch 4 is deactivated below T_{a}, with $\tau = 0.05$, and (b) Patch 4 energy follows the Wentzel-Gunton model with $\Delta s_{\mathrm{w}} = -50$.

Under these two approximations, the crystal chemical potential for the MLG model can be written as

$$\beta\mu_{\mathrm{x}}(\beta, p) \approx \beta_0\mu_{\mathrm{x}}(\beta_0, p) + \frac{p}{\rho}(\beta - \beta_0) + \xi(\beta), \tag{18}$$

where

$$\xi(\beta) = \int_{\beta_0}^{\beta} \Big[-\varepsilon_{\rm tot} + \Delta\varepsilon(1) n_{\rm w} - 2n_w\varepsilon_{\rm w} + 2n_w(\frac{d\Delta s_{\rm w}}{d\beta'} - \beta'\frac{d\varepsilon_{\rm w}}{d\beta'}) \Big] d\beta' \tag{19}$$

$$= (\Delta\varepsilon(1) n_{\rm w} - \varepsilon_{\rm tot})(\beta - \beta_0) - 2n_{\rm w}\int_{\beta_0}^{\beta} \varepsilon_{\rm w} d\beta' + 2n_{\rm w}\Delta s_{\rm w}\Big|_{\beta_0}^{\beta} - 2n_{\rm w}\Big(\beta\varepsilon_{\rm w}\Big|_{\beta_0}^{\beta}\Big) + 2n_{\rm w}\int_{\beta_0}^{\beta} \varepsilon_{\rm w} d\beta' \tag{20}$$

$$= (\Delta\varepsilon(1) n_{\rm w} - \varepsilon_{\rm tot})(\beta - \beta_0) + 2n_{\rm w}\Big[(\Delta s_{\rm w}(\beta) - \Delta s_{\rm w}(\beta_0)) - (\beta\varepsilon_{\rm w}(\beta) - \beta_0\varepsilon_{\rm w}(\beta_0))\Big], \tag{21}$$

and where $-\varepsilon_{\rm tot} = -\sum_{i=1}^{5} \varepsilon_i$ is the temperature independent portion of the crystal ground state energy per particle, i.e., $U_0/N = -\varepsilon_{\rm tot} - n_{\rm w}\Delta\varepsilon(\beta)$, for the MLG model. We thus have

$$\frac{\beta\mu_{\rm x}}{d\beta} \approx (\Delta\varepsilon(1) n_{\rm w} - \varepsilon_{\rm tot}) + 2n_{\rm w}\Big[\frac{d\Delta s_{\rm w}}{d\beta} - \varepsilon_{\rm w} - \beta\frac{d\varepsilon_{\rm w}}{d\beta}\Big], \tag{22}$$

which has a minimum when

$$\Gamma(\beta) \equiv \frac{d\Delta s_{\rm w}}{d\beta} - \varepsilon_{\rm w} - \beta\frac{d\varepsilon_{\rm w}}{d\beta} = \frac{\varepsilon_{\rm tot}}{2n_{\rm w}} - \frac{\Delta\varepsilon(1)}{2}. \tag{23}$$

As noted above, $\beta\mu_{\rm f}$ decreases with decreasing pressure, and because by thermodynamic stability so does ρ, an inverted solubility regime is only obtained when the slope of $\beta\mu_{\rm x}$ with respect to β is positive. For $\Gamma(\beta) > \varepsilon_{\rm tot}/(2n_{\rm w}) - \Delta\varepsilon(1)/2$, the slope of $\beta\mu_{\rm x}$ is positive, hence inverted solubility is observed.

For the Wentzel-Gunton model, the change in $\beta\mu_{\rm x}$ with temperature can be similarly estimated. We can write the energy per particle in the crystal as

$$U_0(\beta)/N = -\varepsilon_{\rm tot} - 2\Delta s_{\rm w}\Big(\frac{1}{\beta_{\rm ref}} - \frac{1}{\beta}\Big). \tag{24}$$

and hence, following Eq. (16),

$$\beta\mu_{\rm x}(\beta, p) = \beta\mu_{\rm x}(\beta_0, p) + \Big(-\varepsilon_{\rm tot} - \frac{2\Delta s_{\rm w}}{\beta_{\rm ref}} - \frac{p}{\rho}\Big)\Big(\beta - \beta_0\Big). \tag{25}$$

The slope of $\beta\mu_{\rm x}$ with respect to β is positive when $-\varepsilon_{\rm tot} - p/\rho > 2\Delta s_{\rm w}/\beta_{\rm ref}$, thus resulting in an inverted solubility regime.

Writing $\beta\mu_{\rm x}$ in a compact form for the temperature-(de)activated patchy model is not possible –the associated integrals need to be evaluated numerically, but the phenomenology is similar. Solubility is inverted in the region where $\beta\mu_{\rm x}$ has a positive slope, i.e., around $T_{\rm a}$, as can be seen in Fig. 4a.

If patch energies are modified by either randomly perturbing them or by scaling them by a constant factor, the free energy of the altered model can be estimated from the original model, assuming that the crystal free energy can be expressed as

$$\beta A' = \beta A - \beta U_0/N + \beta U_0'/N, \tag{26}$$

where A' is the Helmholtz free energy and U_0' is the ground state crystal energy for the altered model. This treatment amounts to neglecting the change in crystal entropy upon weakening or strengthening the patches, which is but a small contribution in this temperature regime. We separately verify that the crystal remains stable at the temperatures of interest.

Put together, various approximations described above allow for the expedited consideration of coexistence points that constitute the solubility curves by generating $\beta\mu_{\rm f}$ and $\beta\mu_{\rm x}$

curves as functions of temperature and pressure.

4 Results and Discussion

4.1 Inverted Solubility from Hydrophobicity Models

In order for the microscopic hydrophobicity models described in Sec. 3.2 to give rise to solubility inversion, a sufficiently large number of water molecules must be involved. In this section we first consider physical bounds on that number, and then consider how the corresponding crystal solubility lines are affected.

4.1.1 Effect of Parameters on Solubility Lines

The key free parameter in hydrophobicity models is the number of water molecules, n_{w}, solvating the hydrophobic patch. We first estimate the number of water molecules potentially available around Patch 4 by calculating the SASA for the participating residues[30] and then computing

$$n_{\mathrm{w}} = A_4 \rho_{\mathrm{w}} \int_{3\text{Å}}^{4.5\text{Å}} g_{\mathrm{C}}(r) dr, \tag{27}$$

where A_4 is the solvent accessible surface area of Patch 4, $\rho_{\mathrm{w}} = 3.3 \times 10^{-2} \text{Å}^{-3}$ is the number density of bulk water at room temperature, and $g_{\mathrm{C}}(r)$ is the radial distribution function of water around carbon atoms determined in Ref. 49. This estimate thus assumes that (i) the solvent has a radius of 1.4Å (the SASA definition), (ii) the average van der Waals radii of protein heavy atoms is ~ 1.6Å, and (iii) the first solvation shell ends with the first peak of $g(r)$ at 4.5Å. We also assume that the measured surface is flat, which is here but a small correction. If we further assume that all residues contributing to Patch 4 are hydrophobic, then $n_{\mathrm{w}} = 133 - 140$ for all six protein structures. However, because Patch 4 contains only a

handful of hydrophobic residues a more realistic estimate should use a smaller A_4. Taking a residue as hydrophobic if it is labeled as such in *any* of the hydrophobicity scales considered in Sec. 2 gives instead $n_{\rm w} = 43 - 48$. Because the hydrophobic residues within Patch 4 are not contiguous, the solvating water molecules are affected by the presence of hydrophilic surface residues nearby. This estimate for $n_{\rm w}$ should thus be treated as an upper bound.

We also consider the number of water molecules needed for Patch 4 to have its measured bonding strength. In particular, if we attribute the entire Patch 4 energy to the change in free energy upon moving solvating water molecules to the bulk, then the MLG model gives $\varepsilon_4 = n_{\rm w}\Delta\varepsilon(1)$, and thus $n_{\rm w} \approx 23$. Because multiple hydrogen bonds also contribute to the patch energy, however, this number should also be treated as an upper bound.

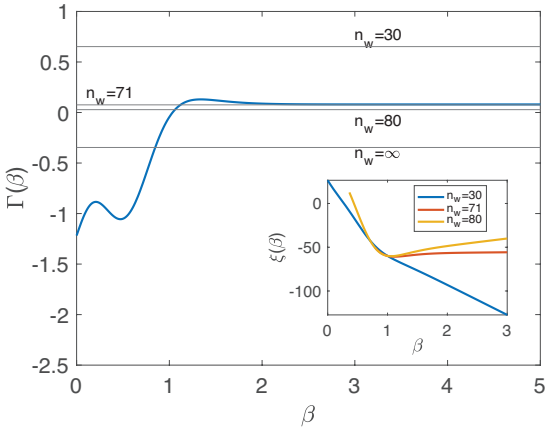

Figure 5: The minimum of $\xi(\beta)$, and hence of $\beta\mu_{\rm x}$, is obtained by the intersection of $n_{\rm w}$ values (black lines) with $\Gamma(\beta)$ (the blue curve) as given in Eq. (23). The inset shows the corresponding $\xi(\beta)$, i.e. the temperature-dependent part of $\beta\mu_{\rm x}$ for various n_w using model parameters reported by Silverstein et al. Here, $n_{\rm w} = 30$ is insufficient to invert solubility, but $n^*_{\rm w} \geq 71$ is.

We can now contrast these bounds with the minimum number of water molecules, $n^*_{\rm w}$, that need to be displaced to invert solubility. For the MLG model, we use Eq. (23) and the sum of DBI patch energies, $\varepsilon_{\rm tot} = 60$, to estimate $n^*_{\rm w}$; it must be such that $\xi(\beta_{\rm min})$ is a minimum, i.e., $\Gamma(\beta) > \varepsilon_{\rm tot}/(2n_{\rm w}) - \Delta\varepsilon(1)/n_{\rm w}$. In other words, solubility is inverted if $\beta > \beta_{\rm min}$. The numerical solution in Fig. 5 shows that $n^*_{\rm w} \gtrsim 71$, and the corresponding change in $\xi(\beta)$ is given in the inset. It should be noted, however, that $n^*_{\rm w}$ depends strongly on the

MLG model parameters. For the multiplicities proposed by Shiryayev et al., for instance, inverted solubility is possible with a mere $n_{\mathrm{w}}^{*} \approx 18$. This second choice of degeneracies, however, seems unphysical;[25] taking $q_{\mathrm{ob}} = 10$ and $q_{\mathrm{os}} = 1$, indeed suggests that the ordered bulk degeneracy is an order of magnitude larger than that of the ordered shell. We thus expect $n_{\mathrm{w}}^{*} \gtrsim 71$ to be a physically more reasonable estimate.

That said, however, because $\varepsilon_{\mathrm{tot}} = 60$ results in room temperature solubilities that are orders of magnitude lower than experimental observations, and in light of the various sources error in patch energy determination (Sec. 3.1), Ref. 30 proposed to halve patch energies in order to reduce the discrepancies. For $\varepsilon_{\mathrm{tot}} = 30$, $n_{\mathrm{w}}^{*} \sim 35$ (Fig. 6), which is less than the 40 or so water molecules solvating hydrophobic residues in Patch 4, but more than the energy-based estimate.

In light of the many estimates involved in the above analysis, the hydrophobic effect as a cause of inverted solubility, although weakly supported, cannot be eliminated outright. Even if the hydrophobicity model parameters are kept constant, a possible resolution could be for Patch 4 to be stronger than estimated and the other patches weaker. The hydrophobicity scenario, however, does severely constrain the patch model parameters. It should further be noted that the P23T mutation is not associated with a systematic change in A_4 (and thus n_{w}), which suggests that a conformational change in the solvating water molecules should accompany the mutation.

4.1.2 Solubility Lines for Models of Hydrophobicity

In Sec. 4.1.1 we determined that the hydrophobicity scenario for inverting solubility requires a fine balance between the protein-protein patch energies, the size of the hydrophobic patch, and the number of water molecules solvating it. While this rare confluence of factors could explain why inverted solubility is not common among proteins, it is natural to wonder whether the presence of weak hydrophobic patches, which are ubiquitous in proteins, affects solubility lines without engendering a regime of inverted solubility. In this section, we study

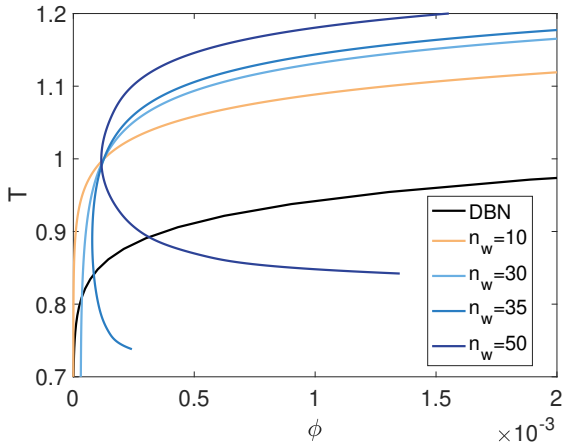

Figure 6: Solubility lines corresponding to different values of n_{w} for $\varepsilon'_{\mathrm{tot}} = 30$, for which $n^*_{\mathrm{w}} \geq 35$.

the Wentzel-Gunton model in order to explore this possibility.

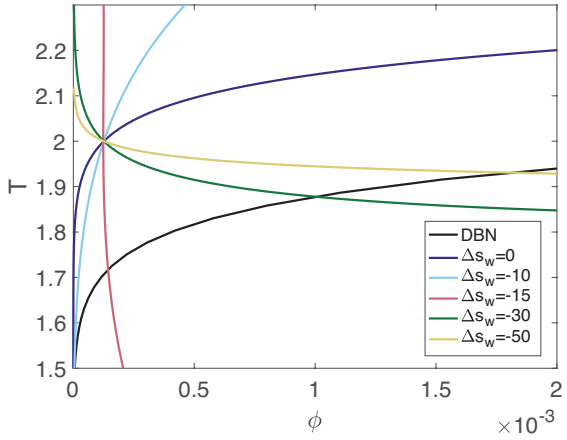

Figure 7: Larger magnitudes of Δs_{w} invert solubility, whereas $\Delta s_{\mathrm{w}} = -15$ results in solubility that only weakly depends on temperature, and $\Delta s_{\mathrm{w}} = -10$ (light blue) results in normal solubility. Note that, for the latter case, even though the solubility is not inverted, the solubility line is markedly altered compared to the $\Delta s_{\mathrm{w}} = 0$ case (dark blue).

The solubility lines for the Wentzel-Gunton model in Fig. 7 are specifically obtained for $\beta_{\mathrm{ref}} = 0.5$, to match the experimental solubility as in Ref. 30, but our observations are qualitatively independent of this choice. Setting $\Delta s_{\mathrm{w}} = -10$, which is here akin to $n_{\mathrm{w}} \approx 20$ (assuming that the temperature-dependent energy in the MLG model scales as n_{w}), results in normal solubility, but the steepness of the solubility curve changes markedly compared

to $\Delta s_{\mathrm{w}} = 0$. Setting $\Delta s_{\mathrm{w}} = -15$ ($n_{\mathrm{w}} \approx 30$) results in the DBI solubility being almost independent of temperature and in DBN being more stable than DBI at $T < T_{\mathrm{tp}} \approx 1.7$. Further reducing Δs_{w} gives rise to an inverted solubility regime. The solubility curve then flattens below $T \approx 2$ and ϕ_{tp} moves to higher packing fractions. These observations thus emphasize that the presence of an inverted solubility regime is the limit case of a continuum of how hydrophobicity impacts the solubility line.

4.2 Solubility Lines for Temperature Deactivated Patches

Absent definitive microscopic evidence for the hydrophobic effect, we finally consider a generic model for patch deactivation. The temperature-deactivated patchy model, which was used to successfully capture the inverted solubility of DBI,[30] stabilizes the crystal with increasing temperature without referring to any specific microscopic mechanism. In this section, we first discuss the physical constraints on the model parameters and then consider how solubility lines change with model parameters, paying particular attention to the robustness of the inverted solubility regime. We also estimate the binodal and the critical temperature, which have been experimentally determined for certain human γD-crystallin mutants.[31]

4.2.1 Parameter Estimates

Despite the absence of an explicit microscopic interpretation for the (de)activation model, one can still place some reasonably solid physical constraints its tuning parameters. First, the (de)activation temperature T_{a} must lie in the vicinity of the triple point, and thus $T_{\mathrm{a}} \approx T_{\mathrm{tp}}$. For our model, the choice $T_{\mathrm{a}} = 1.9$ ensures that the deactivation of Patch 4 makes DBI metastable with respect to DBN for $T < T_{\mathrm{tp}}$. Second, τ, which sets the temperature range over which (de)activation takes place, ought to capture the degree of cooperativity of the underlying microscopic process. It cannot be arbitrarily small, as it would be at a thermodynamic phase transition, because a macroscopic number of components would then need to be involved. It also cannot be arbitrarily large, because the inverted solubility

regime then vanishes. For reference, recall that denaturing a protein takes place over a few degrees, and any smaller scale rearrangement that involves tens to hundreds of atoms should spread over at least $\gtrsim 10K$. We thus here consider a temperature range of $\sim 10K$, which corresponds to setting $\tau = 0.05$.

We first investigate how varying patch energies impacts the phase diagram, keeping $T_a = 1.9$ and $\tau = 0.05$ constant. As previously reported,[30] the resulting phase diagram (Fig. 8a) exhibits a re-entrance regime bounded by the DBI solubility line, as well as a triple point between the fluid and the two crystal forms. The solubility lines that result from perturbing the patch energies by 5% and 10% shift to substantially lower or higher densities, but the existence of an inverted solubility regime is robust. As expected, the errors inherent to the overall parameterization of the model are therefore qualitatively benign.

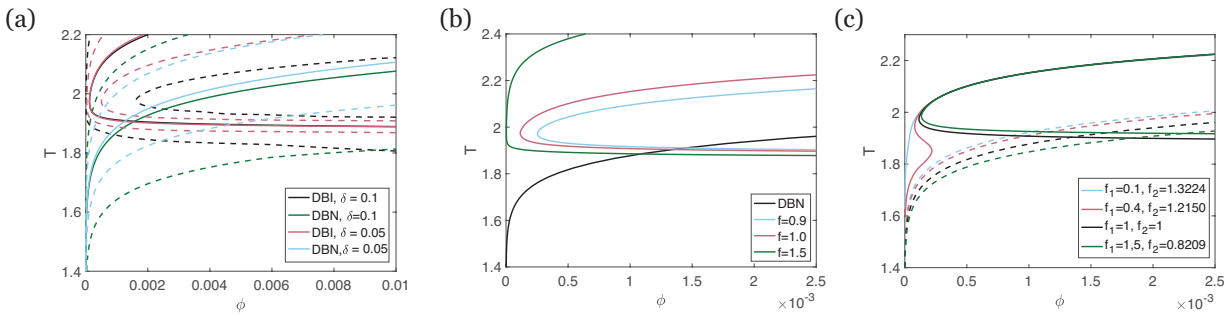

Figure 8: (a) Average solubility lines for perturbed parameters. Dashed lines denote 95% confidence intervals. DBI solubility line for 10% error (black) and 5% error (red), as well as DBN solubility line for 10% error (green) and 5% error (blue) are shown. Higher error levels increase the uncertainty in ϕ_{tp}, as well as the minimum solubility observed for DBI, but inverted solubility is maintained. (b) The effect of changing the energy of the temperature-deactivated patch, such that $\varepsilon_4' = f\varepsilon_4$. (c) The effect of changing ε_4 but keeping the total patch energies of DBI constant.

We then investigate the robustness of the results with respect to the relative strength of the temperature-deactivated fourth patch, ε_4. This question is of interest for two main reasons: (i) the strength and robustness of solubility inversion depend sensitively on the strength of that patch; and (ii) the ordering of the single mutant solubilities directly correlates with their respective Patch 4 energies.

The impact of multiplying ε_4 by $f \in \{0.9, 1.0, 1.5\}$, while keeping the other patch pa-

rameters constant is shown in Fig. 8b. Increasing the strength of Patch 4 systematically decreases the solubility of DBI and lowers ϕ_{tp}. Interestingly, the decrease in solubility with increasing f is consistent with the experimental observations for the single mutants, P23T, P23S, and P23V.[31] Because a stronger Patch 4 decreases the DBI solubility (R36S+P23T double mutant), assuming that the difference between crystals arises due to Patch 4 only, we speculate that if two other double mutants, R36S+P23S and R36S+R23V, were crystallized with similar crystal contacts, then their inverted solubility would order similarly.

The impact of changing the energy of the temperature-dependent patch while keeping the total energy of DBI patches constant, i.e., $\varepsilon_{\mathrm{tot}} = f_1\varepsilon_4 + f_2(\varepsilon_1 + \varepsilon_2 + \varepsilon_3 + \varepsilon_5)$ as shown in Fig. 8c. (Because the second patch corresponds to a shared contact between DBI and DBN, the DBN solubility is then also slightly perturbed.) Here again, the inverted solubility regime vanishes upon markedly reducing the strength of Patch 4. The difference with the first case is that DBN is now metastable with respect to DBI within the probed temperature range, while DBI becomes metastable with respect to DBN otherwise. For $f_1 = 0.4$, DBN is still metastable with respect to DBI, but inverted solubility is observed only over a narrow range of density.

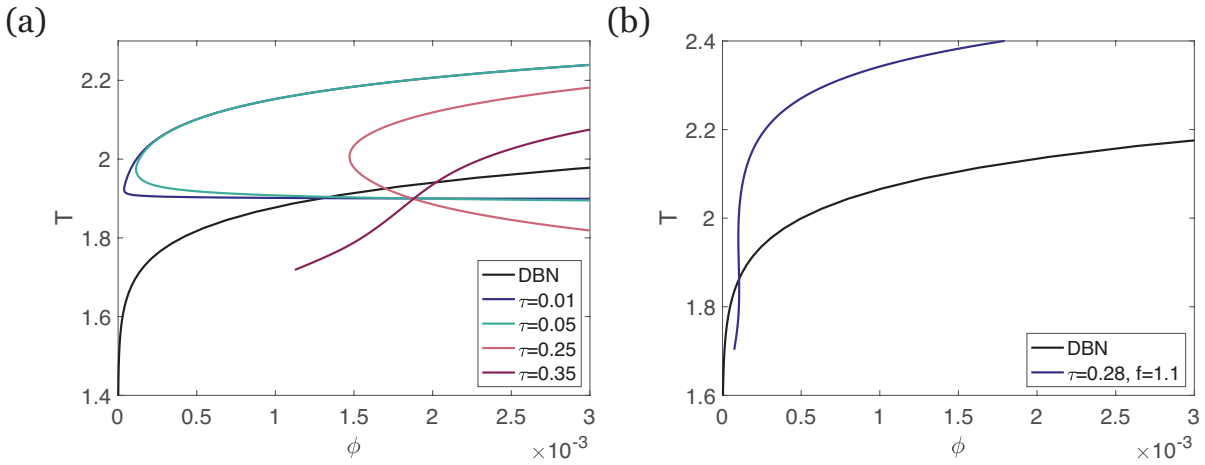

Figure 9: (a) As τ is increased, the DBI solubility line becomes less flat, and eventually inverted solubility is lost (e.g. $\tau = 0.35$). (b) Manipulating the sum of DBI patch energies and τ, one can obtain a temperature range over which the solubility is almost temperature-independent.

We also investigate the robustness of the phenomenology with respect to changes in τ. Decreasing τ corresponds to a faster temperature (de)activation of the patch, which flattens the inverted solubility region and results in $T_{\mathrm{tp}} \to T_{\mathrm{a}}$ as $\tau \to 0$ (Fig. 9a). ϕ_{tp} similarly gets pushed to higher packing fractions, suggesting that a protein solution prepared very near T_{a} could reach remarkably high concentrations compared to solutions prepared at surrounding temperatures. However, as argued above, very small values of τ are physically unreasonable. Conversely, increasing τ weakens this transition and eventually eliminates the inverted solubility regime. Interestingly, a specific choice of τ, with a minor tweak to patch energies ($\tau = 0.28$, $\varepsilon_i' = 1.1\varepsilon_i$), gives rise to a nearly vertical solubility curve (Fig. 9b), similar to the temperature-independent solubility of apoferritin.[21]

4.2.2 Estimation of the Critical Temperature

Although various theoretical results suggest that a closed-loop binodal with multiple critical points is possible upon introducing temperature-dependent binding energies,[23,24,57] no experimental evidence of such a binodal has been found for any human γD-crystallin mutant. In addition, experiments find that the P23V mutation, which also inverts solubility, has a binodal that is indistinguishable from that of the wild type.[31] Here we use Wertheim's perturbation theory,[54,55] which provides quantitatively good estimates of the binodals in patchy models to explore this question (see SI Sec. S4[64]). In order to determine whether our model is consistent with these observations, we estimate the liquid-liquid binodal and the associated critical temperature, T_{c}.

Choosing $\tau = 0.05$ and $T_{\mathrm{a}} = 1.9$, as above, results in a typical binodal with a single critical point at $T_{\mathrm{c}} = 1.85$ (Fig. 10). Hence, without altering patch energies, our model does not give rise to a closed-loop binodal (Fig. 10), consistently with experiments.[31] In order to determine how far our model is from exhibiting a closed-loop binodal, we systematically increase the energy of Patch 4. Only when ε_4 more than doubles does a closed-loop binodal appear. (See Fig. 10). This perturbation, however, falls far outside of the error estimates of

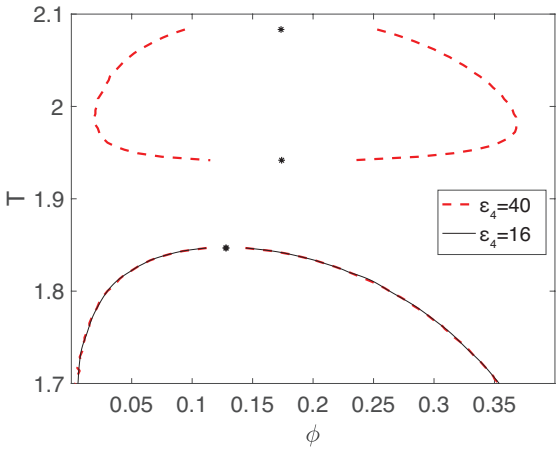

Figure 10: The liquid-liquid binodal regions for the temperature-deactivated patch model for various Patch 4 energies, ε_4. For the reference model patch energy, $\varepsilon_4 = 16$ ($\tau = 0.05$, $T_a = 1.9$), there is no closed-loop binodal. Only if $\varepsilon_4 > 36$, here $\varepsilon_4 = 40$ is shown (red), are multiple critical points and a closed-loop binodal obtained.

the patch energies, which further supports the qualitative robustness of our model prediction.

5 Conclusion

In this article, we have attempted to rationalize the inverted solubility of certain mutants of γD-crystallin based on microscopic models of protein-protein interactions and their temperature dependence. We have paid particular attention to the putative role of hydrophobicity. Estimating surface hydrophobicity using different scales did not reveal the presence of any pertinent surface feature, but microscopic models of hydrophobicity suggest that the number of available surrounding water molecules might suffice. Although our analysis falls short of conclusively determining whether hydrophobicity plays a determining role or not in this protein, this scenario nonetheless seems a bit far fetched due to the lack of structural evidence, as well as the thermodynamic constraints on the patch energies for such a hydrophobic scenario to occur. (By the same token, however, this analysis provides an explanation as to why inverted solubility is far less common than proteins with hydrophobic patches on their surface.) A more conclusive determination would require for the water structure around the

region of interest to be more specifically probed. Because standard water models are insufficiently sensitive to this feature,[49] simulations with more sophisticated water models,[65–68] and neutron diffraction or hydrogen-deuterium exchange experiments might be more productive avenues. For the former, however, extensive testing, including comparison with experimental structures, of the ability of these water models in capturing protein-water interactions would first be needed.

Even though the microscopic origin of inverted solubility in human γD-crystallin still remains somewhat elusive, additional insight from the crystallization of other double mutants, such as R36S+P23V and R36S+P23S, might be helpful in identifying generic features that might have eluded the analysis thus far. Repeating the above structural and thermodynamical for these mutants could help tease out more subtle features that might be at play.

Acknowledgement

PC wishes to thank his colleague David Beratan for his continued support. It has been a pleasure to exchange fruits and science with him over the last decade. PC and IA acknowledge support from National Science Foundation Grant no. NSF DMR-1749374. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562, as well as the Duke Compute Cluster. IA and PC warmly thank Diana Fusco for stimulating discussions. JJM and MKQ were supported by Science Foundation Ireland (grant number 11/RFP.1/PHY/3165). JJM and SJ acknowledge the Irish Research Council for funding. ARK was supported by Science Foundation Ireland (grant number 12/IA/1239). Data relevant to this work have been archived and can be accessed at `https://doi.org/10.7924/XXXXXXX`.

Supporting Information Available

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/XXXX.

Details regarding the analysis of crystal structures, model parameters and geometry, the determination of solubility lines, calculation of the second virial coefficient, estimation of the critical temperature with Wertheim's perturbation theory.

References

(1) McManus, J. J.; Charbonneau, P.; Zaccarelli, E.; Asherie, N. The physics of protein self-assembly. *Curr. Opin. Colloid Interface Sci* **2016**, *22*, 73–79.

(2) Fusco, D.; Charbonneau, P. Soft matter perspective on protein crystal assembly. *Colloids Surf. B* **2016**, *137*, 22–31.

(3) Hagan, M. F. Modeling viral capsid assembly. *Adv. Chem. Phys.* **2014**, *155*, 1.

(4) Šarić, A.; Chebaro, Y. C.; Knowles, T. P.; Frenkel, D. Crucial role of nonspecific interactions in amyloid nucleation. *Proc. Natl. Acad. Sci. U.S.A* **2014**, *111*, 17869–17874.

(5) Suzuki, Y.; Cardone, G.; Restrepo, D.; Zavattieri, P. D.; Baker, T. S.; Tezcan, F. A. Self-assembly of coherently dynamic, auxetic, two-dimensional protein crystals. *Nature* **2016**, *533*, 369.

(6) Pieters, B. J.; van Eldijk, M. B.; Nolte, R. J.; Mecinović, J. Natural supramolecular protein assemblies. *Chem. Soc. Rev.* **2016**, *45*, 24–39.

(7) Glotzer, S. C.; Solomon, M. J. Anisotropy of building blocks and their assembly into complex structures. *Nat. Mater* **2007**, *6*, 557.

(8) Huang, P.-S.; Boyken, S. E.; Baker, D. The coming of age of de novo protein design. *Nature* **2016**, *537*, 320.

(9) Salgado, E. N.; Radford, R. J.; Tezcan, F. A. Metal-directed protein self-assembly. *Acc. Chem. Res* **2010**, *43*, 661–672.

(10) Koehler Leman, J.; Ulmschneider, M. B.; Gray, J. J. Computational modeling of membrane proteins. *Proteins* **2015**, *83*, 1–24.

(11) Altan, I.; Charbonneau, P. *Protein Self-Assembly*; Springer, 2019; pp 209–227.

(12) Vega, C.; Sanz, E.; Abascal, J.; Noya, E. Determination of phase diagrams via computer simulation: methodology and applications to water, electrolytes and proteins. *J. Phys. Condens. Matter* **2008**, *20*, 153101.

(13) Romano, F.; Sanz, E.; Sciortino, F. Phase diagram of a tetrahedral patchy particle model for different interaction ranges. *J. Chem. Phys.* **2010**, *132*, 184501.

(14) ten Wolde, P. R.; Frenkel, D. Enhancement of protein crystal nucleation by critical density fluctuations. *Science* **1997**, *277*, 1975–1978.

(15) Romano, F.; Sanz, E.; Sciortino, F. Role of the range in the fluid- crystal coexistence for a patchy particle model. *J. Phys. Chem. B* **2009**, *113*, 15133–15136.

(16) Lomakin, A.; Asherie, N.; Benedek, G. B. Aeolotopic interactions of globular proteins. *Proc. Natl. Acad. Sci. U.S.A* **1999**, *96*, 9465–9468.

(17) Bianchi, E.; Blaak, R.; Likos, C. N. Patchy colloids: state of the art and perspectives. *Phys. Chem. Chem. Phys* **2011**, *13*, 6397–6410.

(18) McPherson, A. Introduction to protein crystallization. *Methods* **2004**, *34*, 254–265.

(19) Pande, A.; Zhang, J.; Banerjee, P. R.; Puttamadappa, S. S.; Shekhtman, A.; Pande, J. NMR study of the cataract-linked P23T mutant of human γD-crystallin shows minor changes in hydrophobic patches that reflect its retrograde solubility. *Biochem. Biophys. Res. Commun* **2009**, *382*, 196–199.

(20) Pande, A.; Ghosh, K. S.; Banerjee, P. R.; Pande, J. Increase in surface hydrophobicity of the cataract-associated P23T mutant of human γD-crystallin is responsible for its dramatically lower, retrograde solubility. *Biochemistry* **2010**, *49*, 6122–6129.

(21) Vekilov, P. G.; Feeling-Taylor, A.; Yau, S.-T.; Petsev, D. Solvent entropy contribution to the free energy of protein crystallization. *Acta Crystallogr. D* **2002**, *58*, 1611–1616.

(22) Petsev, D. N.; Thomas, B. R.; Yau, S.-T.; Tsekova, D.; Nanev, C.; Wilson, W. W.; Vekilov, P. G. Temperature-independent solubility and interactions between apoferritin monomers and dimers in solution. *J. Cryst. Growth* **2001**, *232*, 21–29.

(23) Shiryayev, A.; Pagan, D. L.; Gunton, J. D.; Rhen, D.; Saxena, A.; Lookman, T. Role of solvent for globular proteins in solution. *J. Chem. Phys.* **2005**, *122*, 234911.

(24) Wentzel, N.; Gunton, J. D. Effect of solvent on the phase diagram of a simple anisotropic model of globular proteins. *J. Phys. Chem. B* **2008**, *112*, 7803–7809.

(25) Moelbert, S.; De Los Rios, P. Hydrophobic interaction model for upper and lower critical solution temperatures. *Macromolecules* **2003**, *36*, 5845–5853.

(26) Lee, B.; Graziano, G. A two-state model of hydrophobic hydration that produces compensating enthalpy and entropy changes. *J. Am. Chem. Soc.* **1996**, *118*, 5163–5168.

(27) Muller, N. Search for a realistic view of hydrophobic effects. *Acc. Chem. Res* **1990**, *23*, 23–28.

(28) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The protein data bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(29) James, S.; Quinn, M. K.; McManus, J. J. The self assembly of proteins; probing patchy protein interactions. *Phys. Chem. Chem. Phys* **2015**, *17*, 5413–5420.

(30) Khan, A. R.; James, S.; Quinn, M. K.; Altan, I.; Charbonneau, P.; McManus, J. J. Temperature-dependent interactions explain normal and inverted solubility in a γD-crystallin mutant. *Biophys. J.* **2019**, *117*, 930–937.

(31) McManus, J. J.; Lomakin, A.; Ogun, O.; Pande, A.; Basan, M.; Pande, J.; Benedek, G. B. Altered phase diagram due to a single point mutation in human γD-crystallin. *Proc. Natl. Acad. Sci. U.S.A* **2007**, *104*, 16856–16861.

(32) Schrödinger, LLC,

(33) Fusco, D.; Barnum, T. J.; Bruno, A. E.; Luft, J. R.; Snell, E. H.; Mukherjee, S.; Charbonneau, P. Statistical analysis of crystallization database links protein physicochemical features with crystallization mechanisms. *PLOS ONE* **2014**, *9*, e101123.

(34) Kyte, J.; Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol* **1982**, *157*, 105–132.

(35) Wimley, W. C.; White, S. H. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat. Struct. Mol. Biol* **1996**, *3*, 842.

(36) Hessa, T.; Kim, H.; Bihlmaier, K.; Lundin, C.; Boekel, J.; Andersson, H.; Nilsson, I.; White, S. H.; von Heijne, G. Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **2005**, *433*, 377.

(37) Moon, C. P.; Fleming, K. G. Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 10174–10177.

(38) Zhao, G.; London, E. An amino acid "transmembrane tendency" scale that approaches the theoretical limit to accuracy for prediction of transmembrane helices: relationship to biological hydrophobicity. *Protein Sci.* **2006**, *15*, 1987–2001.

(39) Basak, A.; Bateman, O.; Slingsby, C.; Pande, A.; Asherie, N.; Ogun, O.; Benedek, G. B.; Pande, J. High-resolution X-ray crystal structures of human γD crystallin (1.25 Å) and the R58H mutant (1.15 Å) associated with aculeiform cataract. *J. Mol. Biol* **2003**, *328*, 1137–1147.

(40) Ji, F.; Koharudin, L. M.; Jung, J.; Gronenborn, A. M. Crystal structure of the cataract-causing P23T γD-crystallin mutant. *Proteins* **2013**, *81*, 1493–1498.

(41) Kmoch, S.; Brynda, J.; Asfaw, B.; Bezouška, K.; Novák, P.; Řezáčová, P.; Ondrová, L.; Filipec, M.; Sedláček, J.; Elleder, M. Link between a novel human γD-crystallin allele and a unique cataract phenotype explained by protein crystallography. *Hum. Mol. Genet.* **2000**, *9*, 1779–1786.

(42) Šali, A.; Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779–815.

(43) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

(44) Fusco, D.; Headd, J. J.; De Simone, A.; Wang, J.; Charbonneau, P. Characterizing protein crystal contacts and their role in crystallization: rubredoxin as a case study. *Soft Matter* **2014**, *10*, 290–302.

(45) Gögelein, C.; Nägele, G.; Tuinier, R.; Gibaud, T.; Stradner, A.; Schurtenberger, P. A simple patchy colloid model for the phase behavior of lysozyme dispersions. *J. Chem. Phys.* **2008**, *129*, 08B615.

(46) Berendsen, H. J.; van der Spoel, D.; van Drunen, R. GROMACS: a message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun* **1995**, *91*, 43–56.

(47) Kästner, J. Umbrella sampling. *Wiley Interdiscip. Rev. Comput. Mol. Sci* **2011**, *1*, 932–942.

(48) Wickstrom, L.; Okur, A.; Simmerling, C. Evaluating the performance of the ff99SB force field based on NMR scalar coupling data. *Biophys. J.* **2009**, *97*, 853–856.

(49) Altan, I.; Fusco, D.; Afonine, P. V.; Charbonneau, P. Learning about Biomolecular Solvation from Water in Protein Crystals. *J. Phys. Chem. B* **2018**, *122*, 2475–2486.

(50) Henzler-Wildman, K.; Kern, D. Dynamic personalities of proteins. *Nature* **2007**, *450*, 964.

(51) Silverstein, K. A.; Haymet, A.; Dill, K. A. Molecular model of hydrophobic solvation. *J. Chem. Phys.* **1999**, *111*, 8000–8009.

(52) Silverstein, K. A.; Haymet, A.; Dill, K. A. A simple model of water and the hydrophobic effect. *J. Am. Chem. Soc.* **1998**, *120*, 3166–3175.

(53) Feyereisen, M. W.; Feller, D.; Dixon, D. A. Hydrogen bond energy of the water dimer. *J. Phys. Chem.* **1996**, *100*, 2993–2997.

(54) Wertheim, M. Fluids with highly directional attractive forces. I. Statistical thermodynamics. *J. Stat. Phys* **1984**, *35*, 19–34.

(55) Wertheim, M. Fluids with highly directional attractive forces. II. Thermodynamic perturbation theory and integral equations. *J. Stat. Phys* **1984**, *35*, 35–47.

(56) Sear, R. P. Phase behavior of a simple model of globular proteins. *J. Chem. Phys.* **1999**, *111*, 4800–4806.

(57) de las Heras, D.; da Gama, M. M. T. Temperature (de) activated patchy colloidal particles. *J. Phys. Condens. Matte* **2016**, *28*, 244008.

(58) Geerts, N.; Eiser, E. DNA-functionalized colloids: Physical properties and applications. *Soft Matter* **2010**, *6*, 4647–4660.

(59) Chen, B.; Siepmann, J. I. A novel Monte Carlo algorithm for simulating strongly associating fluids: Applications to water, hydrogen fluoride, and acetic acid. *J. Phys. Chem. B* **2000**, *104*, 8725–8734.

(60) Bernard, E. P.; Krauth, W.; Wilson, D. B. Event-chain Monte Carlo algorithms for hard-sphere systems. *Phys. Rev. E* **2009**, *80*, 056704.

(61) Kern, N.; Frenkel, D. Fluid–fluid coexistence in colloidal systems with short-ranged strongly directional attraction. *J. Chem. Phys* **2003**, *118*, 9882–9889.

(62) Fusco, D.; Charbonneau, P. Crystallization of asymmetric patchy models for globular proteins in solution. *Phys. Rev. E* **2013**, *88*, 012721.

(63) Frenkel, D.; Ladd, A. J. New Monte Carlo method to compute the free energy of arbitrary solids. Application to the fcc and hcp phases of hard spheres. *J. Chem. Phys* **1984**, *81*, 3188–3193.

(64) Liu, H.; Kumar, S. K.; Sciortino, F.; Evans, G. T. Vapor-liquid coexistence of fluids with attractive patches: An application of Wertheim's theory of association. *J. Chem. Phys.* **2009**, *130*, 044902.

(65) Kessler, J.; Elgabarty, H.; Spura, T.; Karhan, K.; Partovi-Azar, P.; Hassanali, A. A.; Kuhne, T. D. Structure and dynamics of the instantaneous water/vapor interface revisited by path-integral and ab initio molecular dynamics simulations. *J. Phys. Chem. B* **2015**, *119*, 10079–10086.

(66) Tainter, C. J.; Shi, L.; Skinner, J. L. Reparametrized E3B (explicit three-body) water model using the TIP4P/2005 model as a reference. *J. Chem. Theory Comput* **2015**, *11*, 2268–2277.

(67) Levitt, M.; Hirshberg, M.; Sharon, R.; Laidig, K. E.; Daggett, V. Calibration and testing of a water model for simulation of the molecular dynamics of proteins and nucleic acids in solution. *J. Phys. Chem. B* **1997**, *101*, 5051–5061.

(68) Babin, V.; Leforestier, C.; Paesani, F. Development of a "first principles" water potential with flexible monomers: Dimer potential energy surface, VRT spectrum, and second virial coefficient. *J. Chem. Theory Comput* **2013**, *9*, 5395–5403.